



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Matematica e Informatica		
ANNO ACCADEMICO OFFERTA	2021/2022		
ANNO ACCADEMICO EROGAZIONE	2022/2023		
CORSO DILAUREA	INFORMATICA		
INSEGNAMENTO	FONDAMENTI DI SCIENZA DEI DATI		
TIPO DI ATTIVITA'	B		
AMBITO	50166-Discipline Informatiche		
CODICE INSEGNAMENTO	21616		
SETTORI SCIENTIFICO-DISCIPLINARI	INF/01		
DOCENTE RESPONSABILE	GARLISI DOMENICO	Ricercatore a tempo determinato	Univ. di PALERMO
ALTRI DOCENTI			
CFU	6		
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	102		
NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA	48		
PROPEDEUTICITA'			
MUTUAZIONI			
ANNO DI CORSO	2		
PERIODO DELLE LEZIONI	2° semestre		
MODALITA' DI FREQUENZA	Facoltativa		
TIPO DI VALUTAZIONE	Voto in trentesimi		
ORARIO DI RICEVIMENTO DEGLI STUDENTI	GARLISI DOMENICO Mercoledì 15:00 16:30 Stanza 222 presso il Dipartimento di Matematica e Informatica, via Archirafi 34 Venerdì 15:00 16:30 Stanza 222 presso il Dipartimento di Matematica e Informatica, via Archirafi 34		

DOCENTE: Prof. DOMENICO GARLISI

PREREQUISITI	E' necessaria la conoscenza di principi base di analisi, algebra lineare, calcolo delle probabilità, e programmazione.
RISULTATI DI APPRENDIMENTO ATTESI	<p>Conoscenza e capacità di comprensione: Alla fine del corso gli studenti avranno compreso i principali aspetti teorici e pratici della scienza dei dati. Gli studenti conosceranno le varie fasi di un progetto di scienza dei dati e comprenderanno le varie problematiche collegate. Gli studenti avranno preso familiarità con varie tecniche di apprendimento automatico per sfruttare i dati al fine di risolvere compiti complessi.</p> <p>Capacità di applicare conoscenza e comprensione: Gli studenti acquisiranno la capacità di applicare metodi e tecniche per la stima, la riduzione della dimensionalità, la classificazione e il clustering di dati di diversa natura.</p> <p>Autonomia di giudizio: Gli studenti saranno in grado di valutare l'adeguatezza di varie soluzioni algoritmiche per un dato contesto applicativo. Essi saranno in grado di valutare l'efficacia dei vari approcci interpretando opportuni indicatori.</p> <p>Abilità comunicative: Gli studenti acquisiranno la capacità di argomentare e comunicare efficacemente le proprie conclusioni usando il linguaggio della scienza dei dati.</p> <p>Capacità d'apprendimento: Gli studenti acquisiranno la capacità di combinare metodi statistici e computazionali per identificare trend nei dati e generare nuova conoscenza. Essi acquisiranno le competenze necessarie ad apprendere in modo autonomo nuovi metodi e tecniche nell'ambito della scienza dei dati.</p> <p>Competenze acquisite: Gli studenti acquisiranno la capacità di combinare metodi statistici e computazionali per identificare trend nei dati e generare nuova conoscenza.</p>
VALUTAZIONE DELL'APPRENDIMENTO	<p>Le conoscenze e le competenze acquisite dallo studente saranno verificate attraverso una prova scritta e un colloquio orale. Durante il corso, in accordo con il calendario accademico, sarà possibile sostenere una prova scritta in itinere. Tale prova potrà essere completata da una seconda prova scritta parziale svolta in uno degli appelli d'esame.</p> <p>Valutazione della prova scritta La media pesata della prova in itinere e della seconda prova parziale costituisce il voto della prova scritta. In alternativa, lo studente può sostenere una sola prova scritta comprendente tutto il programma del corso. La prova scritta (in itinere, parziale o complessiva) è costituita da esercizi volti a verificare le conoscenze degli argomenti affrontati durante il corso, e di applicare le capacità e le conoscenze acquisite. Nello svolgimento assume fondamentale importanza il commento teorico dei risultati ottenuti. L'articolazione della soluzione consente di apprezzare tutti i livelli di preparazione. La valutazione è espressa in trentesimi ed un minimo di 15 è richiesto per accedere alla prova orale.</p> <p>Valutazione per la prova orale La prova orale consiste in un colloquio, volto ad accertare il possesso delle competenze e delle conoscenze disciplinari previste dal corso; la valutazione viene espressa in trentesimi. Durante il colloquio orale lo studente dovrà essere in grado di discutere le soluzioni proposte durante la prova scritta; inoltre saranno proposte domande di diverso e crescente livello di complessità al fine di valutare il raggiungimento degli obiettivi formativi e le abilità comunicative dello studente. Infine, allo scopo di valutare l'autonomia di giudizio, sarà richiesto di analizzare le caratteristiche di specifici scenari applicativi e di proporre le soluzioni più adeguate ai problemi individuati. La valutazione finale terrà conto sia del punteggio della prova scritta sia dell'esito della prova orale.</p>
OBIETTIVI FORMATIVI	Il corso fornisce competenze riguardanti la scienza dei dati. Lo studente imparerà come trarre conclusioni utili da set di dati ampi e diversificati attraverso l'esplorazione, la previsione e l'inferenza. Gli studenti impareranno a porre domande appropriate sui propri dati e a interpretare correttamente le risposte fornite dagli strumenti inferenziali e computazionali. Lo studente prenderà familiarità con tecniche di apprendimento automatico per sfruttare i dati al fine di risolvere compiti complessi. Le esercitazioni comprendono l'implementazione di algoritmi in Python. A tale scopo, una parte del corso sarà dedicata a fornire agli studenti la conoscenza di base tale linguaggio.
ORGANIZZAZIONE DELLA DIDATTICA	Le lezioni sono organizzate utilizzando metodologie di didattica interattiva, attività di apprendimento attivo, e didattica capovolta.

TESTI CONSIGLIATI	<p>- Testi principali / main textbooks [dimensionality reduction, machine learning] :</p> <p>Blum, Hopcroft, and Kannan, Foundations of Data Science, Cambridge University Press, 2020; ISBN: 978-1108485067</p> <p>Brunton and Kutz, Data Driven Science & Engineering: Machine Learning, Dynamical Systems, and Control, Cambridge University Press, 2019. ISBN: 978-1108422093</p> <p>- Testi per consultazione / reference books [Python, data analysis]:</p> <p>Sinan Ozdemir, Data Science: guida ai principi e alle tecniche base della scienza dei dati, Apogeo. ISBN: 8850334192</p> <p>Sebastian Raschka, Machine learning con python. Apogeo. ISBN: 9788850333974</p>
--------------------------	--

PROGRAMMA

ORE	Lezioni
4	Python: Programmare con numeri e stringhe. Decisioni. Cicli. Funzioni.
4	Python: Liste, tuple, insiemi e dizionari. Librerie NumPy e Scikit-learn.
2	Le fasi di un progetto di scienza dei dati: Determinare gli obiettivi. Ottenere i dati. Analisi esplorativa dei dati. Il problema della qualità dei dati: pre-elaborazione e pulizia dei dati. Costruire e validare un modello.
6	Analisi esplorativa dei dati - Capire i dati: ricerca di relazioni tra dati, verifica di appartenenza a distribuzioni, ricerca di outlier, standardizzazione, covarianza.
2	Costruire buoni set di addestramento: la pre-elaborazione - La qualità dei dati. Rimozione e imputazione dei valori mancanti da un dataset. Miglioramento della collocazione in categorie dei dati per gli algoritmi di machine learning. Selezione delle caratteristiche rilevanti per la costruzione di un modello.
4	Riduzione della dimensionalità: La maledizione della dimensionalità. Singular Value Decomposition (SVD).
4	Riduzione della dimensionalità: Approssimazione di matrici ed esempi sulla compressione delle immagini. Principal Component Analysis (PCA)
2	Introduzione al machine learning: Tipi di machine learning. Metodi supervisionati e non supervisionati. Regressione, classificazione, clustering.
4	Apprendimento supervisionato: Classificatore bayesiano, classificatore k-vicini
4	Apprendimento supervisionato: alberi decisionali
4	Apprendimento non supervisionato: Introduzione al clustering, similarità e dissimilarità, errore quadratico, algoritmi di clustering gerarchici: single link, complete link, average link
4	Apprendimento non supervisionato: algoritmi di clustering partizionali: kmeans
4	Valutazione dei modelli e ottimizzazione degli iperparametri: Convalida incrociata k-fold. Metodo del "gomito". Metriche per la valutazione delle prestazioni: precision, recall, f-measure. Grafici ROC